



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension

Citation for published version:

Nieuwland, MS & Martin, AE 2012, 'If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension', *Cognition*, vol. 122, no. 1, pp. 102-109.
<https://doi.org/10.1016/j.cognition.2011.09.001>

Digital Object Identifier (DOI):

[10.1016/j.cognition.2011.09.001](https://doi.org/10.1016/j.cognition.2011.09.001)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Cognition

Publisher Rights Statement:

© Nieuwland, M. S. & Martin, A. E. Jan 2012, If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension, in *Cognition*. 122, 1, p. 102-109 © Elsevier. The final publication is available at <http://dx.doi.org/10.1016/j.cognition.2011.09.001>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



If the real world were irrelevant, so to speak: The role of propositional truth-value in
counterfactual sentence comprehension

Mante S. Nieuwland & Andrea E. Martin

Basque Center on Cognition, Brain and Language

Word count: 4030

Date: September 9th, 2011

Keywords: Language comprehension, real-world knowledge, propositional truth-value,
contextual relevance, counterfactual conditionals, Event-Related Potentials, N400

Address correspondence to:

Mante S. Nieuwland

Basque Center on Cognition, Brain and Language (BCBL)

Paseo Mikeletegi 69-2, 20009, Donostia-San Sebastian, Spain

ACKNOWLEDGEMENTS

We are grateful to Javi Miqueleiz for help with stimulus construction and data collection, and to Heather Ferguson and two anonymous reviewers for helpful comments on an earlier version of this manuscript. This work is supported by a Plan Nacional research grant from the Spanish Ministry of Science and Innovation (PSI2010-18087).

ABSTRACT

Propositional truth-value can be a defining feature of a sentence's relevance to the unfolding discourse, and establishing propositional truth-value in context can be key to successful interpretation. In the current study, we investigate its role in the comprehension of counterfactual conditionals, which describe imaginary consequences of hypothetical events, and are thought to require keeping in mind both what is true and what is false. Pre-stored real-world knowledge may therefore intrude upon and delay counterfactual comprehension, which is predicted by some accounts of discourse comprehension, and has been observed during online comprehension. The impact of propositional truth-value may thus be delayed in counterfactual conditionals, as also claimed for sentences containing other types of logical operators (e.g., negation, scalar quantifiers). In an Event-Related Potential (ERP) experiment, we investigated the impact of propositional truth-value when described consequences are both true and predictable given the counterfactual premise. False words elicited larger N400 ERPs than true words, in negated counterfactual sentences (e.g., "If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been Russia/America") and real-world sentences (e.g., "Because N.A.S.A. developed its Apollo Project, the first country to land on the moon was America/Russia") alike. These indistinguishable N400 effects of propositional truth-value, elicited by opposite word pairs, argue against disruptions by real-world knowledge during counterfactual comprehension, and suggest that incoming words are mapped onto the counterfactual context without any delay. Thus, provided a sufficiently constraining context, propositional truth-value rapidly impacts ongoing semantic processing, be the proposition factual or counterfactual.

INTRODUCTION

What if N.A.S.A. had never developed its Apollo Project? Would the Soviet Union have ‘won’ the Space Race? Would Neil Armstrong ever have coined his famous words strolling across the moonscape? Counterfactual reasoning, our ability to consider this hypothetical scenario, is pervasive in everyday life (e.g., Byrne, 2002; Kahneman & Miller, 1986; Roese, 1997), and considered one of the hallmarks of complex reasoning skills (e.g., Braine & O’Brien, 1991; Johnson-Laird & Byrne, 2002). The decoupling from reality in language and thought may reflect an important step in human evolution that greatly boosted cognitive and communicative power (e.g., Cosmides & Tooby, 2000), but also presents complex processing challenges. After all, if language use draws upon facts and fiction, how do we keep the two apart?

Counterfactuals provide a unique opportunity to study how people establish propositional truth-value during language comprehension, the routine processes upon we rely daily to determine whether what we hear or read is consistent with what we already know. Propositional truth-value, although essential in linguistic and philosophical theories of meaning (e.g., Montague, 1973; Tarski, 1944), has not had the centrality in psycholinguistic research that its import suggests. This is striking, because propositional truth-value is a defining feature, although clearly not the only one, of a sentence’s relevance to the unfolding discourse (e.g., Wilson & Sperber, 2002). Thus, establishing propositional truth-value in context can be key to successful interpretation. Research on the processing of truth-value has focused on logical operators (e.g., negation, scalar quantifiers), and its results have been taken to suggest that propositional truth-value does not modulate early measures of semantic processing (e.g., Fischler, Bloom, Childers, Roucos, & Perry, 1983;

Kounios & Holcomb, 1992). However, this field of research has often ignored the role of contextual relevance: what is literally true need not be a prototypically relevant or plausible thing to say (e.g., that a robin is not a tree), and early effects of propositional truth-value have been observed using more naturalistic materials (e.g., Nieuwland & Kuperberg, 2008). In the current study, we investigate the role of propositional truth-value during comprehension of negated counterfactuals about historical events. Such sentences entail a unique challenge to the language comprehension system, because following a counterfactual premise (e.g., If N.A.S.A. had never developed its Apollo Project), described consequences that are consistent with current facts of the real world (e.g., the first country to land on the moon being the U.S.A.) may become false, and, vice versa, consequences that are in fact false (e.g., the first country to land on the moon being the Soviet Union) may become hypothetically true.

The relationship between fiction and reality has been central in philosophy of language and cognitive science (e.g., Gerrig & Prentice, 1991; Searle, 1975). Extant theories of human cognition, such as the Conceptual Blending framework (e.g., Coulson, 2001; Turner & Fauconnier, 1998) and Mental Models theory (e.g., Byrne & Johnson-Laird, 2009), sometimes assume that conflicting representations (e.g., of what is true and of what is false) are simultaneously active (see also de Vega, Urrutia & Rizzo, 2007). However, while counterfactual thoughts are commonly expressed through language, such theories don't specify how this representational conflict plays out during language processing. In contrast, 'memory-based' language processing theories (e.g., Gerrig & O'Brien, 2005; Kintsch, 1988; Myers & O'Brien, 1998; Sanford & Garrod, 2005) posit that words initially activate pre-stored world knowledge and earlier concepts from the text, and

that the contents of active memory are subsequently integrated into the discourse context by inhibiting contextually irrelevant concepts. Importantly, because the initial stage is blind to contextual relevance or propositional truth-value, world knowledge could hinder ongoing comprehension. Note that this may particularly be true in counterfactual sentences: if factual information is held active during comprehension (e.g., Byrne, 2002), this information may act as a ‘lure’ that makes it harder to falsify a counterfactual consequence (e.g., the first to land on the moon would be the U.S.A.).

Recent results support effects of pre-existing real-world knowledge on counterfactual comprehension. Ferguson and Sanford (2008) showed that sentences describing implausible real-world events (e.g., “Families would feed their cat a bowl of carrots”) incurred temporary disruptions during sentence reading, as indexed by longer early fixations in eye-tracking, despite a counterfactual context (e.g., “If cats were vegetarians”; see Ferguson, Scheepers & Sanford, 2010, for evidence from visual world eye-tracking). Ferguson, Sanford and Leuthold (2008) replicated this result with negated counterfactuals (e.g., “If cats were not carnivores...”) and showed that counterfactual context did not preclude implausible events from eliciting an enlarged N400 ERP, a negative voltage deflection in the electroencephalogram that indexes early semantic processing (Kutas & Hillyard, 1980). Importantly, this N400 effect aligned with the first-pass reading results, whereas later eye-tracking measures suggested that propositions were eventually accommodated into the context. These early disruptions as indexed by first-pass reading times and, crucially, the N400, were taken as evidence that mapping utterances onto pre-existing world knowledge is an inevitable consequence of the memory retrieval mechanisms by which we compute meaning, and that counterfactual context comes into

play at a later moment (Ferguson & Sanford, 2008; Ferguson et al., 2008; but see Ferguson, Sanford & Leuthold, 2007, who reported N400 evidence for rapid integration into an affirmative, rather than negated counterfactual world). These conclusions resonate with the long-held assumption that non-propositional semantic processes precede post-semantic decision processes that compute sentence truth-value, based on the insensitivity of the N400 to propositional truth-value in sentences containing logical operators (e.g., Fischler et al., 1983; Kounios & Holcomb, 1992; Urbach & Kutas, 2010; but see Nieuwland, Ditman & Kuperberg, 2010; Nieuwland & Kuperberg, 2008).

Although an account wherein counterfactual truth-value comes into play relatively late (but not necessarily after lower-level processes are completed) captures extant results, it seems hard to reconcile with well-established rapid effects of various types of context on the processing of simple declarative sentences (e.g., Altmann & Kamide, 1999; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Van Berkum, 2009). Perhaps due to the semantically complex nature of counterfactuals (e.g., Kratzer, 1989), incremental build-up of sentence meaning proceeds slowly, thereby delaying any impact of propositional truth-value, similar to what has been claimed for negation and scalar quantifiers (e.g., Fischler et al., 1983). However, an alternative explanation for a delayed impact takes into account whether the context is indeed sufficiently supportive for the unfolding message. In possible world semantics, the correct interpretation of an unfolding counterfactual resembles the actual world as closely as possible with the exception of what is explicitly or implicitly required (e.g., Lewis, 1973; Stalnaker, 1968). For example, “If cats were vegetarians” makes it less *implausible*, but not necessarily plausible or true, that cat-owners would feed their cats carrots, whereas it clearly presupposes that cats would not eat meat. In contrast,

“If cats enjoyed eating root vegetables” creates a context in which feeding cats carrots fits well. Thus, rationality in counterfactual thinking imposes relevance constraints (e.g., Evans, 2006), and observed real-world interference may in fact reflect low counterfactual relevance. Importantly, successful communication requires linguistic information to be relevant to and coherent with the discourse (e.g., Wilson & Sperber, 2002), whether or not its contents abide by the physical and biological laws and historical facts of our particular world (e.g., Stalnaker, 1968). Eye-tracking and ERP studies on story comprehension have shown that contextual relevance can outweigh real-world plausibility from an early moment on (e.g., “the peanut was in love” is processed more easily than “the peanut was salted” following a cartoon-like story about an amorous peanut; e.g., Filik, 2008; Nieuwland & Van Berkum, 2006), consistent with an account wherein extra-linguistic information is rapidly assimilated into a rich mental representation of the relevant context, paving the way for retrieving the next word’s meaning, sometimes even before the next word has been encountered in the unfolding utterance (e.g., Kutas, Van Petten, & Kluender, 2006; Otten, Nieuwland & Van Berkum, 2007; Van Berkum, 2009). However, when contextual constraints are weak, early eye-tracking measures and the N400 seem to be driven also by low-level lexical-semantic factors (i.e., ‘associative priming’ via lexical-associative or categorical relationships, as indexed by norms of association, relatedness or semantic categories; e.g., Camblin, Gordon & Swaab, 2007; Federmeier & Kutas, 1999; Otten & Van Berkum, 2007). Crucially, therefore, the validity of real-world interference effects stands or falls with the provision of a sufficiently constraining context, so that described

consequences are true and relatively predictable given the counterfactual premise¹. In the current study, this was ensured by the inclusion of two pre-tests.

The current study addressed the role of counterfactual truth-value during sentence comprehension using ERPs, which provide quantitative and qualitative information well in advance of (and without the principled need for) explicit behavioral responses. Our hypothesis focused on the N400, an ERP waveform whose amplitude peaks at about 400 ms post-stimulus, with smaller amplitudes indexing facilitated retrieval from semantic memory as elicited by content words or other meaningful stimuli (Kutas et al., 2006). N400 effect onset, when ERP waveforms corresponding to different conditions start to diverge, is about 200-300 ms after visual word onset. With spoken words, N400 effects start as early as 150-200 ms after word onset, after having heard only two or three phonemes and well before a word's uniqueness point (e.g., Van Petten, Coulson, Rubin, Plante, & Parks, 1999). These well-established observations suggest that N400 effects elicited by less predictable words or semantic anomalies reflect routine, early sense-making processes by which every incoming word is related to the preceding context. In the current study, we evaluated whether these processes are sensitive to propositional truth-value in a negated counterfactual context, or whether they are primarily driven by pre-existing world knowledge.

¹ Plausibility ratings collected by Ferguson and colleagues (Ferguson, personal communication) suggest that counterfactually consistent continuations were indeed less plausible than real-world continuations (3.6 and 4.6, respectively, out of 5-‘highly plausible’). Note that Warren, McConnell and Rayner (2009) also reported increased early fixations for impossible sentences despite a fantasy context, but the provided context was also relatively weak, as reflected in low naturalness ratings (2.69 out of 5-‘very unnatural’) and lower cloze completion scores for impossible sentences than for possible sentences following the fantasy context (7% and 32%, respectively).

Participants read counterfactual and real-world sentences containing critical words belonging to word pairs (e.g., ‘Russia’ and ‘America’). These pairs were chosen so that one word rendered the counterfactual sentence true (e.g., “If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been Russia”), whereas the other rendered the real-world sentence true (e.g., “Because N.A.S.A. developed its Apollo Project, the first country to land on the moon was America”), and so that each word was predictable from the context (see Table 1). False sentences were formed by exchanging words from each word pair. **Note that for counterfactual and real-world sentences alike, establishing whether the sentence is true or false requires real-world knowledge (e.g., knowledge of the ‘Space Race’ between Russia and America which culminated in America’s Apollo Project).** If automatically activated real-world knowledge delays the impact of propositional truth-value, then critical words in counterfactual true sentences and real-world false sentences should both evoke larger N400s than those in counterfactual false sentences and real-world true sentences. An alternative version of this hypothesis is that real-world knowledge and propositional truth-value form simultaneous constraints, so that the effect of truth-value is reduced in counterfactual sentences compared to real-world sentences. In contrast, if propositional truth-value impacts semantic processing without delay, false sentences should elicit larger N400s than true sentences, for counterfactual and real-world sentences alike. Critically, we predicted this result in face of the fact that the N400 effects of propositional truth-value are elicited by opposite pairs of critical words.

Table 1. Example sentences and approximate translations with average truth-value rating and cloze value of the critical word for each condition			
Condition	Example sentences	Mean rating	Mean cloze

		of truth- value	value (%)
Counterfactual True (CT)	<i>Si la N.A.S.A. no hubiera desarrollado su proyecto Apollo, el primer país en pisar la luna habría sido <u>Rusia</u> seguramente.</i> “If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been <u>Russia</u> surely.”	5.62 (.97)	68 (22)
Counterfactual False (CF)	<i>Si la N.A.S.A. no hubiera desarrollado su proyecto Apollo, el primer país en pisar la luna habría sido <u>América</u> seguramente.</i> “If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been <u>America</u> surely.”	1.68 (.74)	-
Real-World True (RWT)	<i>Como la N.A.S.A. desarrolló su proyecto Apollo, el primer país en pisar la luna ha sido <u>América</u> seguramente.</i> “Because N.A.S.A. developed its Apollo Project, the first country to land on the moon was <u>America</u> surely.”	5.50 (1.03)	65 (25)
Real-World False (RWF)	<i>Como la N.A.S.A. desarrolló su proyecto Apollo, el primer país en pisar la luna ha sido <u>Rusia</u> seguramente.</i> “Because N.A.S.A. developed its Apollo Project, the first country to land on the moon was <u>Russia</u> surely.”	1.55 (.64)	-
Note. Standard deviations are given in parentheses. Critical words are underlined for expository purposes. For truth-value rating, 1 = False, 7 = True.			

METHOD

Development and Pretest of Materials

We constructed 133 Spanish sentence quadruplets with two counterfactual and two real-world sentences (see Table 1). Critical words were predicates, nouns or proper names, and never sentence-final. Counterfactual sentences described hypothetical consequences of common-knowledge historical events not having taken place, whereas real-world sentences

described actual consequences of these events. The two sentence types differed in three respects: counterfactuals started with the conditional ‘Si’, contained a negative premise, and involved conditional verb tense, real-world sentences started with ‘Como’ (‘because’), were affirmative, and contained no conditional verb tense.

We first established the expectedness of critical words. Twenty students of the University of the Basque Country completed one of two lists with one version of each item truncated before the critical word. They were instructed to complete the sentence with the first sensible word coming to mind. Cloze value was computed as the percentage of participants who used the intended critical word.

We subsequently determined whether sentences (truncated after the critical word) were, on average, regarded as true or false. Twenty-four different students evaluated one of four counterbalanced sentence lists containing only one condition per quadruplet, and decided whether the sentences were true (1 = False, 7 = True), skipping any they could not evaluate.

Based on these results, we excluded quadruplets with low cloze value, containing true/false sentences rated below/over 4, or sentences skipped by more than two participants. In the ultimate set of 120 quadruplets, true and false sentences had similar cloze values and ratings across conditions (see Table 1), and critical words were matched for mean log frequency (CT/RWT = 1.44/1.55; $p = .11$; Davis & Perea, 2005) and word length (CT/RWT = 6.65/6.89 letters; $p = .24$).

For the ERP experiment, we created four counterbalanced lists so that each sentence appeared in only one condition per list, but in all conditions equally often across lists. We

also included 60 filler sentences, which did not start with ‘Si’ or ‘Como’, and consisted of two clauses separated by a comma.

ERP Experiment

Participants

Thirty right-handed students (12 males; mean age = 21.3 years) gave written informed consent. All were healthy, native Spanish speakers who had not participated in the pretests.

Procedure

ERP participants read sentences from a monitor (black letters, bright background). All sentences were preceded by a fixation cross upon which participants could start the next sentence by button-press. The first clause of each sentence was presented for 4000 ms, followed by a fixation cross and blank screen each for 500 ms; the second clause was presented word-by-word (400 ms word duration, 200 ms inter-word-interval).

Sentence-final words were sometimes followed by a yes/no comprehension question that probed world knowledge related to the sentence. These 60 questions (30 requiring a ‘yes’ button-press response, participants performed at 91 % accuracy on average) were distributed across sentence types. Participants completed a practice-session and six break-separated experimental sessions. Total time-on-task was 50 minutes.

Electroencephalogram Recording

The electroencephalogram (EEG) was recorded from 29 scalp electrodes (left mastoid reference; 1 additional right mastoid electrode and 4 electrooculogram electrodes), amplified (band-pass filtered at 0.01–30 Hz), and digitized at 250 Hz. Impedance was kept

below 5 kOhm. We corrected ocular artifacts using independent component analysis, and single-trial waveforms were automatically screened for artifacts during 1000-ms epochs (starting 250 ms before critical word onset). Three participants were excluded due to excessive artifacts (trial loss > 40 %). For the remaining 27 participants, average ERPs (normalized by subtraction to a 250-ms pre-stimulus baseline) were computed over artifact-free trials per condition (average trial loss = 4.4 %).

RESULTS

Critical words elicited larger (more negative) N400s in the counterfactual false and real-world false sentences compared to counterfactual true and real-world true sentences (see Figure 1). N400 effects in counterfactual and real-world sentences started at about 250 ms after critical word onset, dissipated before 600 ms, and had a broad central-posterior distribution. We performed a 2(factuality: counterfactual, real-world) \times 2(truth-value: true, false) repeated measures analysis of variance (ANOVA) using average amplitude across all EEG electrodes in 4 consecutive 100 ms time windows starting at 200 ms after critical word onset (see Table 2). False sentences elicited larger N400s than true sentences in time windows between 300 and 500 ms. Crucially, there was no significant main effect of factuality ($F < 1.1$ for factuality in all time windows), and the effect of truth-value did not differ for counterfactual and real-world sentences in any time window ($F < 1$ for all factuality by truth-value interactions).

We divided the electrodes into those anterior (FP1/2, FZ, F3/4, F7/8, FC1/2, FC5/6) or posterior (CP1/2, CP5/6, PZ, P3/4, P7/8, O1/2) to the central cross-line. Subsequent 2(factuality: counterfactual, real-world) \times 2(truth-value: true, false) \times 2(distribution:

anterior, posterior) repeated measures ANOVAs showed that false sentences elicit larger N400 differences at posterior electrodes than at anterior electrodes in the 300-400 window ($F(1, 26) = 10.12, p_{(rep)} = .97, \eta_p^2 = .28$) and the 400-500 window ($F(1, 26) = 5.44, p_{(rep)} = .912, \eta_p^2 = .173$), consistent with the typical posterior N400 distribution. This distribution was similar for N400 effects elicited by counterfactual and real-world sentences ($F < 1$ for all 3-way interactions). Follow-up ANOVAs for the posterior and anterior electrode selection separately showed that false sentences elicited larger N400s at anterior electrodes only in the 300-400 ms time window, but at posterior electrodes in all time windows between 300-500 ms (see Table 2).

Table 2. Electrophysiological effect of sentence truth-value (false minus true) across counterfactual sentences and real-world sentences. Presented are average voltage difference (mean M with standard deviation SD in parenthesis), and F value and effect size (partial eta square η_p^2) results from ANOVAs using mean amplitude across all EEG electrodes and in anterior and posterior electrode selections, in 4 consecutive 100 ms time windows starting at 200 ms after critical word onset

	Electrodes		200-300	300-400	400-500	500-600
Truth-value (False minus True)	All	μV	-.42 (2.59)	-1.8 (2.47)	-1.68 (3.31)	-.43 (3.15)
		F	.72	14.7***	6.94*	.49
		η_p^2	.03	.36	.21	.02
	Anterior	μV	-.37 (3.02)	-1.29 (2.87)	-1.13 (3.63)	-.12 (3.56)

		F	.40	5.46*	2.61	.03
		η_p^2	.02	.17	.09	.001
	Posterior	μV	-.50 (2.65)	-2.28 (2.55)	-2.12 (3.27)	1.18
		F	.97	21.63***	11.38**	.29
		η_p^2	.04	.45	.30	.04
For all F tests, numerator $df = 1$, denominator $df = 26$. * = $p \leq .05$, $p_{(rep)} \geq .88$. ** = $p \leq .01$, $p_{(rep)} \geq .95$. *** = $p \leq .001$, $p_{(rep)} \geq .986$.						

Additional 2(factuality: counterfactual, real-world) \times 2(truth-value: true, false) \times 2(hemisphere: left, right) repeated measures ANOVAs showed that the effect of truth-value did not differ across hemispheres in any of the time windows ($F < 1$ for all truth-value by hemisphere interactions) and that this bilateral distribution was similar for counterfactual and real-world sentences ($F < 1$ for all 3-way interactions).

DISCUSSION

We examined neural activity while participants read counterfactual and real-world sentences varying in truth-value. Counterfactual false sentences elicited larger N400s than counterfactual true sentences, and this N400 effect of propositional truth-value was indistinguishable from that in real-world sentences. Our results argue against disruptions by automatically activated real-world knowledge during counterfactual comprehension, and

are at odds with a delayed contribution of context (e.g. Kintsch, 1988; Myers & O'Brien, 1998; Sanford & Garrod, 2005). Instead, they suggest that incoming words can be mapped without any delay onto the most relevant interpretive context, in this case, a counterfactual world. Pre-existing factual knowledge did not reflexively impede counterfactual comprehension, at least to the extent that counterfactual and real-world false sentences similarly disrupted semantic processing, and that counterfactual true sentences did not incur semantic processing costs over real-world true sentences. This is principally evidenced by the inverted N400 responses to the same lexical items as a function of propositional truth-value. More generally, the results are consistent with theories of language comprehension that do not assume a temporal divide between discourse context and world knowledge (see Cook & Myers, 2004). The results provide further evidence that people effortlessly map incoming utterances onto what they think is true and what they consider relevant (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004; Nieuwland & Kuperberg, 2008; Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009), and may reflect how contextual constraints guide expectations about upcoming input (e.g., DeLong, Urbach, & Kutas, 2005; Van Berkum, 2009). We take our results as evidence for relatively coarse-grained anticipation, a background of expectations of relevance that are revised or elaborated as sentences unfold (e.g., Altmann & Mirković, 2009; Van Berkum, 2009, for discussion).

Our results appear to be at odds with those reported by Ferguson and colleagues (Ferguson & Sanford, 2008; Ferguson et al., 2008), who showed that counterfactual context did not impact first-pass reading times or the N400 in a subsequent target sentence. We believe that this apparent contradiction can be accounted for by differences in the strength

of contextual constraints². Our study involved counterfactual and real-world sentences that were considered to be equally true, containing critical words that were similarly predictable. In the studies by Ferguson and colleagues, propositions may have been congruent with the counterfactual context, but perhaps did not follow plausibly or predictably from it. Although we cannot exclude that their results directly reflect the effects of plausibility, their results may reflect low-level lexical-associative factors (e.g., ‘cat-feed-carrots’ versus ‘cat-feed-fish’) that dominate early measures of processing when contextual constraints are weak (e.g., Camblin et al., 2007; Otten & Van Berkum, 2007), rather than reflecting an inherent aspect of the language processing architecture. We wish to note that our own results need not be contingent on the relatively high predictability of the critical words, but we do assume they are contingent on having counterfactual and real-world sentences that are matched on contextual constraint leading up to the critical word and propositional truth-value as rendered by the critical word.

Another factor that may have contributed to the divergence between our results and those of Ferguson colleagues could be the type of counterfactual materials that were used. Alternative endings to known historical events may be relatively easily computed, for example because relevant information is also part of our real-world knowledge (e.g., of the ‘Space Race’ and the fact that the Soviets were also making substantial progress in landing somebody on the moon at the time that the USA managed to do so), in contrast to counterfactual worlds that need to be construed from novel and unfamiliar content (such as

² Differences in contextual constraint may have been exacerbated by the fact that our manipulation occurred in a counterfactual condition sentence whereas earlier studies on counterfactual comprehension have looked at later effects (but see Stewart, Haigh & Kidd, 2009). However, we do not know of any principled reason why there should be differences between counterfactual sentence and discourse manipulations.

the materials used by Ferguson and colleagues). Theories of counterfactual comprehension indeed assume that similarity between counterfactual worlds and the real world facilitates counterfactual reasoning (e.g., Lewis, 1973; see Byrne, 2007, for discussion).

Earlier reported evidence that incoming words are automatically checked against real-world knowledge (Ferguson & Sanford, 2008; Ferguson et al., 2008) has been taken to support the notion that people entertain incompatible mental models during counterfactual comprehension (Byrne & Johnson-Laird, 2009; see also de Vega, Urrutia, & Rizzo, 2007). Although our results do not speak directly to this debate, this mapping onto real-world knowledge may index low counterfactual or contextual relevance rather than the presence of competing mental models. The fact that even the most relevant distracter, arguably receiving strongest lexical-associative priming by the preceding context (e.g., ‘America’ following ‘N.A.S.A.-Apollo Project-land-moon’), had no visible effect on N400 amplitude suggests that subjects only considered the most relevant, counterfactual situation (e.g., Evans, 2006). However, competing mental models do not necessarily or perhaps sufficiently impact retrieval of word-elicited semantic knowledge during online processing to modulate N400 amplitude. On a related note, we do not wish to claim that counterfactual comprehension is identical to real-world sentence comprehension, especially at other moments in the sentence, such as the construction of the counterfactual context. Rather, we propose that at the moment that critical words were encountered, implications of the counterfactual context and real-world context had both been computed such that propositional truth-value directly impacted semantic processing. Importantly, the N400 effects of propositional truth-value in the counterfactual sentences and in the real-world sentences were elicited by opposite pairs of critical words (see Nieuwland & Kuperberg,

2008, for a similar result using affirmative and negative sentences) suggesting that effects associated with particular lexical items can be completely overruled in a strong discourse context.

To conclude, if described consequences are true and predictable given the counterfactual premise, real-world knowledge does not impede counterfactual comprehension, at least at the moment that propositional truth-value can be established. Thus, provided a sufficiently constraining context, propositional truth-value rapidly impacts ongoing semantic processing, be the proposition factual or counterfactual.

LITERATURE

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G.T.M. & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583-609.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of *If*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203.
- Byrne, R.M.J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, 6(10), 426-431.
- Byrne, R.M.J. (2007). Précis of the rational imagination: How people create alternatives to reality. *Behavioral and Brain Sciences*, 30, 439–480.
- Byrne, R.M.J. & Johnson-Laird, P.N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, 13(7), 282-287.
- Camblin, C.C., Gordon, P.C., Swaab, T.Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, 56, 103–128.
- Cook, A., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, 50, 268–288.
- Cosmides, L, & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In Sperber, D. (Ed.), *Metarepresentations* (pp. 53-115). Oxford: Oxford University Press.
- Coulson, S. (2001). *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge: Cambridge University Press.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665-671.
- Delong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.

- de Vega, M., Urrutia, M., & Rizzo, R. (2007). Cancelling updating in the comprehension of counterfactuals embedded in narratives. *Memory & Cognition*, 35, 1410-1421.
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, 13(3), 378-395.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469-95.
- Ferguson, H.J., & Sanford, A.J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, 58, 609-626.
- Ferguson, H.J., Sanford, A.J. & Leuthold, H. (2007). Real-world interference in detecting violations of counterfactual and negated worlds. Poster presented at Cognitive Neuroscience Society Annual meeting. New York, NY.
- Ferguson, H.J., Sanford, A.J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time-course of processing negation and remitting counterfactual worlds. *Brain Research*, 1236, 113-125.
- Ferguson H.J., Scheepers C. & Sanford A.J. (2010). Expectations in counterfactual and theory of mind reasoning, *Language and Cognitive Processes*, 25(3), 297-346.
- Filik, R., (2008). Contextual override of pragmatic anomalies: evidence from eye movements. *Cognition*, 106(2), 1038-46.
- Fischler, I., Bloom, P., Childers, D., Roucos, S., & Perry, N. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400–409.
- Gerrig, R.J. & O'Brien, E.J. (2005). The Scope of Memory-Based Processing. *Discourse Processes*, 39(2-3), 225-242
- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, 2, 336–340.
- Hagoort, P. Hald, L., Bastiaansen, M., & Petersson, K-M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.
- Johnson-Laird, P.N., & Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics and inference. *Psychological Review*, 109, 646-78.
- Kahneman, D., & Miller, D. T. (1986). Norm theory—Comparing reality to its alternatives. *Psychological Review*, 93, 136–153.

- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kratzer, A. (1989). An investigation of the lumps of thought. *Linguistics and Philosophy*, 12, 607–653.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.
- Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.
- Lewis, D. (1973). *Counterfactuals*. Cambridge, Mass.: Harvard University Press.
- Montague, R., (1974). The Proper Treatment of Quantification in Ordinary English. In Thomason R. (ed.), *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven.
- Myers, J.L. & O'Brien, E.J., (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131-157.
- Nieuwland, M.S., Ditman, T. & Kuperberg, G.R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63: 324-346.
- Nieuwland, M.S. & Kuperberg, G.R. (2008). When the truth isn't too hard to handle: An Event-Related Potential study on the Pragmatics of Negation. *Psychological Science*, 19 (12), 1213-1218.
- Nieuwland, M.S. & Van Berkum, J.J.A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098-1111.
- Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, 8, 89.
- Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153, 166-177.
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121, 133–148.

- Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes*, 39, 205–223.
- Searle, J.R. (1975). The logical status of fictional discourse. *New Literary History*, 6, 319-332.
- Stalnaker, R.C. (1968). A theory of conditionals. *American Philosophical Quarterly Monograph Series* (pp. 98–112), N. Rescher (ed.). Blackwell, Oxford.
- Stewart, A.J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Quarterly Journal of Experimental Psychology*, 62, 2113-2125.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268 (5217), 1632-1634.
- Tarski, A. (1944). The Semantic Conception of Truth. In Martinich, A.P. (ed.), pages 69-91. *The Philosophy of Language*. New York: Oxford University Press
- Turner, M. & Fauconnier, G. (1998). Conceptual Integration in Counterfactuals. In Koenig, J.-P. (Ed.). *Conceptual Structure, Discourse, and Language, II*, Stanford: Center for the Study of Language and Information.
- Urbach, T.P., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63, 158-179.
- Van Berkum, J.J.A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.
- Van Berkum, J. J. A., Holleman, B., Nieuwland, M. S., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, 20, 1092 -1099.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 394–417.

- Warren, T., McConnell, K. & Rayner, K. (2008). Effects of context on eye movements when reading about plausible and impossible events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(4), 1001-1010.
- Wilson, D. & Sperber, D. (2002). Truthfulness and relevance, *Mind*, 111, 583-632.

FIGURE CAPTIONS

Figure 1. Electrophysiological effects of truth-value in counterfactual sentences and real-world sentences. The waveforms show the grand average event-related potentials (ERPs) elicited by critical words per condition at 8 selected electrodes. Negative voltage is plotted upwards and waveforms are filtered (5 Hz high cut-off, 12 dB/oct) for presentation purpose only. Stimuli consisted of counterfactual and real-world sentences that were either true or false (examples are provided above the graphs, critical words are underlined). Scalp distributions of the difference effects (false minus true sentences) in adjoining 100 ms analysis windows are given below the graphs.

Counterfactual True

If the N.A.S.A had not developed its Apollo Project, the first country to land on the moon would be Russia...

Counterfactual False

If the N.A.S.A had not developed its Apollo Project, the first country to land on the moon would be America...

Real-world True

Because the N.A.S.A developed its Apollo Project, the first country to land on the moon has been America...

Real-world False

Because the N.A.S.A developed its Apollo Project, the first country to land on the moon has been Russia...

